Construction d'une ontologie pour la prise en charge de l'hypertension artérielle

Olivier Steichen, Christel Daniel-Le Bozec, Marie-Christine Jaulent, Jean Charlet

Construction d'une ontologie pour la prise en charge de l'hypertension artérielle

Olivier Steichen[1], Christel Daniel-Le Bozec[1,2], Marie-Christine Jaulent[1], Jean Charlet[1,3]

[1] INSERM UMR_S 872, Eq. 20, 75006 Paris ;

[2] AP-HP, Hôpital Européen Georges Pompidou, DIH, 75015 Paris ;

[3] AP-HP, STIM/DSI, 75014 Paris.

ost@club-internet.fr, {jean.charlet,christel.lebozec, marie-christine.jaulent}@spim.jussieu.fr

Résumé : Une ontologie pour la prise en charge de l'hypertension est construite pour assister l'analyse qualitative des décisions médicales dans un service spécialisé en hypertension artérielle. La modélisation s'appuie sur trois sources complémentaires. La structure de l'observation médicale informatisée utilisée dans l'unité fournit directement des concepts fondamentaux pour la prise en charge de l'hypertension. L'analyse terminologique de deux corpus permet ensuite d'enrichir et de structurer ce noyau ontologique. Le premier corpus comprend huit guides de bonne pratique qui définissent la prise en charge idéale des patients hypertendus. Le second collige les commentaires en texte libre issus des observations médicales saisies sur une année. Ces commentaires rendent compte des situations effectivement rencontrées en pratique clinique. La combinaison des sources terminologiques est nécessaire pour couvrir le spectre complet des décisions médicales et de leurs justifications dans le contexte d'étude.

Mots-clés : Ontologies, dossier médical informatisé, guides de bonne pratique, évaluation des pratiques.

1 Introduction

1.1 Évaluation des pratiques professionnelles

Les guides de bonne pratique synthétisent dans une forme utilisable par les cliniciens les connaissances les mieux étayées sur une question médicale. Ils sont principalement destinés à favoriser l'application des résultats de la recherche clinique mais servent également de référence pour l'évaluation des pratiques professionnelles.

Cette seconde utilisation soulève trois problèmes principaux. Premièrement, certaines situations rencontrées en médecine spécialisée relèvent bien d'une prise en charge réglée par les résultats de la recherche clinique mais se situent hors de la portée des

recommandations, destinées principalement aux non-spécialistes. Dans ce cas de figure, il serait possible d'expliciter des règles supplémentaires pour compléter les recommandations comme référence pour l'évaluation des pratiques dans un service spécialisé. Deuxièmement, la prise en charge de certaines situations n'a pas été étudiée par la recherche clinique. En l'absence de données suffisantes pour établir des règles, les décisions sont nécessairement individualisées. Enfin, les résultats de la recherche clinique ne sont jamais les seuls déterminants des décisions médicales, même lorsqu'ils permettent de définir des règles de prise en charge. Les caractéristiques des patients, les spécificités de leurs problèmes et les circonstances particulières de prise en charge doivent également être intégrées dans la décision (Haynes *et al.*, 2002). Ces particularités justifient parfois d'agir en désaccord avec les recommandations des guides de bonne pratique ou les règles de prise en charge spécialisée.

1.2 Méthodes qualitatives et besoins ontologiques

Même individualisées en l'absence de règles de prise en charge ou en désaccord avec les règles existantes, les décisions doivent rester justifiées. Leur évaluation fait alors appel aux méthodes qualitatives, qui s'appuient sur l'analyse détaillée et comparative des cas individuels (Green & Britten, 1998; Friedman & Wyatt, 2006). Au vu des caractéristiques particulières d'un patient et d'une situation clinique, il est possible de comprendre les justifications des décisions individualisées.

Dans le contexte d'une unité hospitalière spécialisée en hypertension artérielle, nous souhaitons étudier les déterminants et la logique de l'individualisation des pratiques. Ceci n'est possible à grande échelle qu'avec l'aide d'outils informatiques, qui peuvent notamment regrouper des patients similaires et faciliter l'analyse comparative des décisions médicales. Au préalable, les cas doivent être représentés formellement et avec suffisamment de détail pour le niveau d'analyse visé.

Les ontologies médicales sont des systèmes terminologiques riches et formalisés, développés pour le partage et l'exploitation du contenu sémantique des dossiers médicaux (Charlet *et al.*, 2006). Dans la mesure où chaque ontologie possède une couverture, une granularité et une structure qui lui est propre, elle ne peut habituellement pas être réutilisée pour une autre tâche que celle prévue initialement (Coiera, 1995). Notre objectif actuel est donc de construire une ontologie dédiée à la prise en charge de l'hypertension, en vue d'assister l'analyse qualitative des décisions médicales dans un service spécialisé.

1.3 Ressources pour la modélisation ontologique

La construction des ontologies est classiquement divisée en une phase de repérage des concepts pertinents et une phase d'organisation de ces concepts, à l'aide de relations plus ou moins nombreuses et complexes. Les ressources utilisables pour repérer les concepts d'un domaine et leurs interrelations appartiennent à deux grandes classes (Steichen *et al.*, 2006). La première comprend les conceptualisations théoriques faites par les experts de leur domaine (pour la médecine clinique : formulaires structurés d'observation, guides de bonne pratique, ouvrages académiques, etc.). L'identification des

concepts est plus ou moins directe selon le degré de formalisation de la ressource et la connaissance préalable du domaine par l'ingénieur des connaissances. Lorsque la ressource est très structurée, avec une bonne connaissance du domaine il est possible d'identifier directement les concepts sous-jacents. Lorsque la ressource est en texte libre peu structuré, les outils de traitement automatique des langues peuvent fournir une aide précieuse pour l'identification des concepts.

La seconde classe de ressources utiles pour la modélisation ontologique regroupe les traces empiriques laissées sous forme textuelle par les experts lors de la réalisation de leur activité (pour la médecine clinique : observations médicales, comptes rendus d'hospitalisation, courriers, etc.). En effet, cette production textuelle reflète les concepts de bas niveau dont les experts ont besoin en pratique quotidienne. Évidemment, ces textes ne contiennent pas les concepts de haut niveau qui structurent la représentation cognitive du domaine. Néanmoins, ces concepts structurants peuvent, dans une certaine mesure, être inférés des rapprochements entre des concepts de bas niveau opérés par une analyse terminologique. L'extraction des concepts et l'analyse de leurs interrelations nécessite le recours à de gros corpus représentatifs de l'activité et donc à des outils de traitement automatique des langues pour leur analyse.

Nous avons exploité des ressources des deux types : conceptualisations théoriques de la prise en charge de l'hypertension et traces textuelle concrètes laissées lors de la pratique clinique correspondante.

2 Matériel et méthode

2.1 Matériel

Trois ressources ont été utilisées pour trouver les concepts utiles à la description et à la justification des décisions médicales durant la prise en charge de l'hypertension :

1. les items du formulaire d'observation médicale utilisé dans l'unité d'hypertension, qui permet la saisie structurée des déterminants de la pratique réglée en milieu spécialisé ;
2. un corpus de guides de bonne pratique, qui comportent les déterminants de la pratique réglée en médecine générale ;
3. un corpus de commentaires en texte libre issus d'observations médicales, qui permettent aux médecins spécialistes de consigner (i) les déterminants de la pratique réglée absents de la partie structurée du formulaire et (ii) les déterminants de la pratique individualisée.

Formulaire d'observation. Un formulaire informatisé est utilisé depuis 30 ans dans l'unité d'hypertension pour recueillir les observations médicales. Environ 5 000 observations semi-structurées sont saisies chaque année.

Le modèle d'information clinique initial a été conceptualisé par des experts du domaine. Il a évolué en fonction du progrès des connaissances médicales et du retour d'expérience sur son utilisation. Sa validité est attestée (i) par la possibilité de prendre en charge les patients en s'appuyant uniquement sur l'observation informatisée, (ii) par

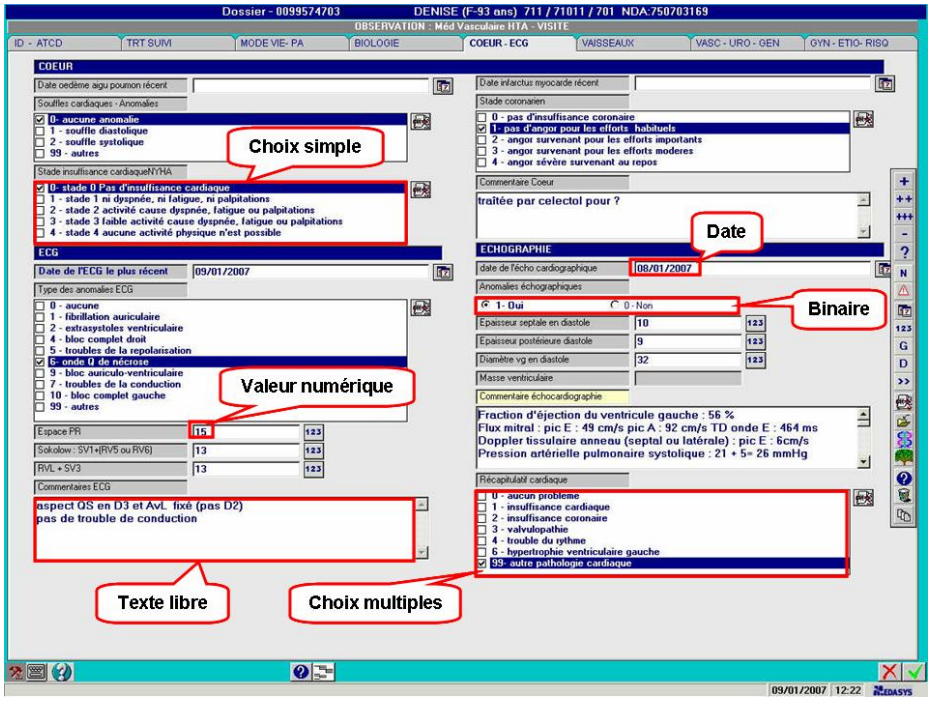

FIG. 1 – Onglet « coeur - ECG » de l'observation médicale.

la possibilité de faire fonctionner des système d'aide à la décision efficaces à partir des données recueillies et (iii) par les nombreux travaux rétrospectifs de recherche clinique auxquels les données recueillies ont pu donner lieu (Degoulet *et al.*, 1990).

Le formulaire compte 176 questions, pour la plupart optionnelles (FIG. 1). Elles sont à réponses booléennes (diabète ? oui/non), temporelles (date du dernier infarctus du myocarde ? 21/12/1996), prédéfinies dans une liste à choix unique ou multiple (récapitulatif cardiaque ? avec une liste de choix comportant, entre autres, « hypertrophie ventriculaire gauche » et « valvulopathie »), numériques (pression artérielle diastolique en millimètres de mercure ? 95) ou textuelles libres (conclusion ? prévoir des explorations hormonales en raison d'une hypokaliémie persistant à l'arrêt des diurétiques). Les réponses sont enregistrées dans une base de données relationnelle. La liste des items cliniques trouvés dans le formulaire (en-têtes de sections et de sous-sections, questions, réponses prédéfinies) a été extraite de la base de données hospitalière à l'aide d'une requête SQL.

Commentaires en texte libre. Les réponses en texte libre des 5 109 observations saisies en 2005 ont été anonymisées : les noms des patients et des médecins ont été remplacés par des identifiants numériques et les dates de naissance par l'âge en années. Leur collection forme un corpus de 350 000 mots en langue française.

Guides de bonne pratique. Guides de bonne pratique. Plutôt que de nous appuyer sur un seul guide et d'en extraire manuellement les concepts, nous avons choisi d'en colliger huit, publiés entre 1999 et 2005, pour constituer un corpus de 56 000 mots en langue anglaise susceptible d'être analysé à l'aide d'outils de traitement automatique des langues. Ce choix repose sur le souhait de traiter de façon homogène les documents en langue naturelle. Chaque guide de bonne pratique étant peu redondant, nous en avons colligé plusieurs afin d'augmenter la puissance des analyses statistiques.

2.2 Méthode

Démarche de construction. Le formulaire d'observation et les guides de bonne pratique sont des conceptualisations théoriques de la prise en charge de l'hypertension artérielle, élaborées par des experts informés de l'état des connaissances médicales dans leur domaine. Les deux premières étapes de modélisation s'appuient sur ces ressources, qui fournissent les concepts clé intervenant dans la prise en charge réglée des hypertendus et des indices pour structurer l'ontologie.

Les commentaires en texte libre saisis dans les observations comportent quelques déterminants d'une pratique réglée absents des items de l'observation. Ils donnent surtout accès aux problèmes concrets et imprévus rencontrés par les cliniciens dans leur pratique quotidienne. Ils alimentent la dernière étape de notre démarche de modélisation, destinée à enrichir le noyau ontologique précédent et éventuellement à corriger sa structure, avec les concepts utilisés pour la description et la justification des décisions individualisées.

A chaque étape de modélisation, les concepts sont liés manuellement aux termes de la SNOMED-CT (version 01/2006) (Wang *et al.*, 2002) en s'aidant du navigateur SNOMED développé à l'université de Virginia (Virginia-Maryland Regional College of Veterinary Medicine, 2007). Ces liens font correspondre notre ontologie de domaine avec une ontologie de référence. Ils fournissent également l'opportunité d'évaluer la couverture du domaine par la SNOMED-CT et la pertinence de sa structure.

Traitement du langage naturel. Les outils de traitement automatique du langage sont nécessaires pour analyser les gros corpus. SYNTEX et UPÉRY, choisis pour l'analyse des commentaires en texte libre et des guides de bonne pratique, peuvent travailler sur des corpus en langue française ou anglaise (Bourigault, 2002; Charlet *et al.*, 2006) .

À l'issue d'un prétraitement morphosyntaxique du corpus, ils identifient les candidats termes susceptibles de représenter des concepts pertinents du domaine, et comptent leurs occurrences. Ils procèdent ensuite à une analyse distributionnelle des occurrences de termes. Les termes qui se présentent régulièrement dans des contextes lexicaux similaires sont alors regroupés. En effet, ce point commun lexical peut indiquer une communauté sémantique utile pour identifier les synonymes et surtout pour structurer l'ontologie. Ainsi, « sténose de l'artère rénale » (SAR) et « hyperaldostéronisme » (HA) partagent de nombreux contextes d'occurrence dans les commentaires en texte libre : « les investigations ont révélé un(e) SAR/HA », « trouver des arguments pour un(e) SAR/HA », « SAR/HA démontré(e) », « absence de SAR/HA », etc. De fait, SAR et HA partagent ces contextes d'occurrence parce qu'ils représentent tous les deux des causes d'hypertension artérielle.

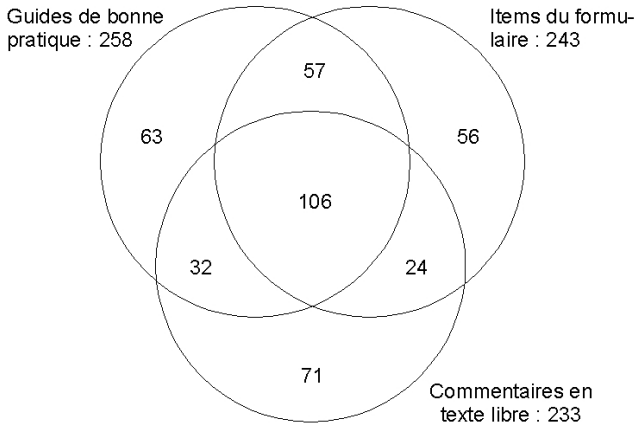

FIG. 2 – Nombre de concepts issus de chaque source (409 concepts au total).

Édition de l'ontologie. La modélisation ontologique est réalisée, concept par concept, dans l'éditeur DOE (Differential Ontology Editor) (Troncy, 2006). Cet outil permet de renseigner, pour chaque concept, les quatre principes issus de la sémantique différentielle qui justifient sa position dans l'ontologie, à savoir communauté et différence avec le père et les frères. La structure obtenue de cette manière est une monohiérarchie strictement taxinomique.

3 Résultats

3.1 État d'avancement

Les concepts provenant des items du formulaire ont été repérés. Les guides de bonne pratique ont été analysés et les concepts correspondants identifiés. Dans un premier temps, seuls les concepts rencontrés plus de 50 fois dans le corpus des commentaires en texte libre ont été extraits. Ce seuil correspond à environ 1% des questionnaires et a été choisi comme compromis entre rappel et bruit pour trouver les concepts récurrents liés à des aspects génériques de la prise en charge des patients hypertendus, à l'instar des concepts issus des items du formulaire et des guides de bonne pratique. Ainsi, nous disposons des concepts fondamentaux utilisés pour la prise en charge réglée de l'hypertension en milieu spécialisé. La FIG. 2 montre que chacune des sources de concepts a contribué significativement à la constitution de ce noyau ontologique. Pour l'heure, seuls les concepts provenant des items du formulaire ont été liés à des termes de la SNOMED-CT.

3.2 Exploitation des conceptualisations expertes du domaine

Concepts issus des items du formulaire d'observation. Les 176 questions du formulaire et les 177 réponses prédéfinies des questions à choix simples ou multiples ont directement fourni 243 concepts cliniques utilisés pour la prise en charge des patients hypertendus. Un même concept, qualifié ou coordonné différemment, peut sous-tendre plusieurs items. Ainsi le concept d'infarctus du myocarde est lié à la question « date de l'infarctus du myocarde le plus récent » mais aussi à une des réponses prédéfinies à la question à choix multiples « récapitulatif cardiaque ». La pertinence de ces concepts est garantie par la validité clinique et scientifique du formulaire. Un terme correspondant a été trouvé dans la SNOMED-CT pour 212 d'entre eux.

Au-delà des concepts cliniques provenant des questions et des réponses prédéfinies, les titres des sections et des sous-sections de l'observation correspondent à des concepts structurants, comme celui d'histoire familiale, et révèlent comment les cliniciens se figurent leur domaine. Ces concepts structurants ont été notés comme indices pour la structuration ontologique.

Concepts issus des guides de bonne pratique. L'analyse des guides de bonne pratique a révélé 258 concepts cliniques : 163 déjà identifiés dans les items du formulaire et 95 concepts additionnels.

Les 80 concepts provenant des items du questionnaire non retrouvés dans les guides de bonne pratique sont très spécialisés (par exemple l'infarctus rénal comme cause d'hypertension), de granularité très fine (les glomérulopathies comme sous classe des maladies rénales) ou réfèrent à la prise en charge spécifique d'autres facteurs de risque cardiovasculaire (comme le diabète et son traitement). Deux exemples de concepts additionnels parmi les 95 trouvés dans les guides de bonne pratique sont celui d'apnées du sommeil comme cause de l'hypertension et celui de démence comme conséquence de l'hypertension. De par la nature des guides de bonne pratique, la pertinence de ces concepts pour la prise en charge non spécialisée de l'hypertension est validée par la recherche clinique.

L'analyse terminologique des guides de bonne pratique a également révélé des concepts structurants, comme celui d'atteinte des organes cibles. L'analyse distributionnelle n'a pas été très productive sur ce corpus peu redondant. Elle a tout de même fait apparaître quelques groupes sémantiques intéressants. Différentes caractéristiques de l'hypertension ont ainsi été rapprochées et peuvent s'organiser selon de nombreux axes : hypertension permanente, épisodique, blouse-blanche ou masquée (selon les moments où l'hypertension est présente) ; hypertension globale ou systolique isolée (selon la ou les composante(s) élevée(s) de la pression artérielle) ; hypertension artérielle légère, modérée ou sévère (selon le niveau de pression artérielle) ; etc.

3.3 Exploitation des textes produits en pratique courante

L'analyse des commentaires en texte libre identifie 233 concepts fréquents (ayant plus de 50 occurrences dans le corpus). Seulement 130 des 243 concepts issus des items du formulaire sont retrouvés parmi eux. En effet, si un concept est déjà pris en compte dans la partie structurée de l'observation, les médecins n'ont plus besoin de le réutiliser dans

les commentaires en texte libre, sauf parfois pour apporter des précisions. Trente deux des 103 autres concepts trouvés dans les commentaires en texte libre figurent également dans les guides de bonne pratique et 71 sont complètement originaux.

Validation des concepts issus des guides de bonne pratique. Les concepts issus des guides de bonne pratique sont pertinents pour la pratique en milieu spécialisé s'ils sont effectivement utilisés pour la prise en charge des patients de l'unité. Dans ce cas, ou bien ils se trouvent dans la partie structurée de l'observation (163 concepts mentionnés dans la section précédente), ou bien ils sont utilisés par les cliniciens dans leurs commentaires en texte libre (au moins les 32 autres concepts mentionnés dans le paragraphe précédent, plus ceux qui apparaissent moins de 50 fois dans le corpus que nous ne pouvons pas encore quantifier).

Ainsi, les commentaires en texte libre comptent 89 occurrences de termes référant au concept d'apnées du sommeil, comme « apnées du sommeil » ou « ronchopathie ». La pertinence en milieu spécialisé de ce concept trouvé dans les guides de bonne pratique est donc validée par son emploi effectif dans la prise en charge des patients de l'unité. Par opposition, aucune occurrence du terme « démence » n'est trouvée dans les commentaires en texte libre. Les termes connexes, comme « troubles de la mémoire », sont également rares (moins de 10 occurrences pour l'ensemble). En pratique, la question des troubles cognitifs ne se pose pas dans l'unité spécialisée, car les patients en question ne sont pas adressés en consultation d'hypertension par leurs généralistes. Néanmoins, il paraît difficile de ne pas retenir ce concept dans l'ontologie, compte-tenu de son importance théorique et pratique dans une vision globale de la prise en charge des patients hypertendus.

Nouveaux concepts. L'analyse du corpus de commentaires en texte libre permet également d'identifier 71 concepts absents des guides de bonne pratique mais fréquemment utilisés par les cliniciens pour décrire ou justifier leurs décisions.

La plupart de ces concepts s'intègrent dans des protocoles de prise en charge spécialisée, adaptés au recrutement de l'unité d'hypertension (forte prévalence des hypertensions compliquées et/ou avec une cause spécifique). Par opposition, les guides de bonne pratique sont surtout destinés à la prise en charge des hypertensions non compliquées et idiopathiques en médecine générale. Le concept de dissociation rénine – aldostérone, par exemple, compte 80 occurrences dans les commentaires en texte libre et constitue un déterminant important pour la prise en charge de certains patients, selon des règles de décision ou des habitudes locales. Ce concept ne relève donc pas de l'individualisation des pratiques mais d'un autre niveau de prise en charge réglée, au-delà des recommandations trouvées dans les guides de bonne pratique.

Néanmoins, parmi ces concepts fréquemment rencontrés dans les commentaires en texte libre, on en trouve déjà qui se rapportent aux particularités de certains cas réclamant une individualisation des décisions. Ainsi, la notion de refus du patient (d'une hospitalisation, d'un traitement, du sevrage tabagique, de l'inclusion dans un protocole, etc.) n'est pas explicitée dans les guides de bonne pratique et ne relève d'aucune règle de prise en charge. Cependant, elle apparaît régulièrement dans les commentaires en texte libre et influence significativement les décisions médicales.

Enfin, ce corpus de commentaires en texte libre indique lui aussi des principes d'organisation conceptuelle, cette fois-ci plus par l'analyse distributionnelle que par la

présence de concepts structurants. Par exemple, de nombreux concepts liés aux décisions thérapeutiques médicamenteuses ont été regroupés : ajuster ou simplifier le traitement, débuter un nouveau médicament, continuer ou interrompre un ancien médicament, augmenter ou diminuer une posologie, etc.

4 Discussion

4.1 Nécessité d'une ontologie spécifique

Comme nous l'avons évoqué en introduction, une ontologie contribuant à une tâche particulière dans un domaine restreint requiert un pouvoir expressif suffisant (étendue et niveau de granularité de la conceptualisation) et un potentiel calculatoire adéquat (lié à la structure ontologique). Par conséquent, les ontologies générales comme la SNOMED-CT peuvent difficilement être réutilisées pour des applications particulières imprévues. Leur intérêt comme ontologies de référence pour une interopérabilité sémantique n'est évidemment pas remis en question. Le travail de mise en correspondance des concepts du noyau ontologique avec la SNOMED-CT a confirmé les limites de cette dernière qui condamnent une réutilisation directe pour notre projet.

Tout d'abord, établir une correspondance univoque n'est pas toujours possible car certains termes de la SNOMED-CT sont ambigus. Le concept de glomérulopathie, trouvé dans l'observation parmi les maladies rénales, peut être mis en correspondance avec deux termes distincts de la SNOMED-CT : « renal glomerular disease » [code 76910007] et « glomerular disease » [197679002]. La distinction entre ces termes est difficile à saisir, d'autant plus que deux de leurs descendants respectifs sont sans discussion possible des synonymes stricts : « malignancy associated glomerulonephritis » [236508005] et « glomerular disorders in neoplastic diseases » [197738008].

Alors que la SNOMED-CT comporte plus de 308 000 termes, sa couverture du domaine est imparfaite. Même lorsqu'on ne considère que les 243 concepts issus des items du formulaire, 31 (13%) n'ont pas de correspondant dans la SNOMED-CT. Ainsi, il n'y a pas de terme référant aux hypertensions monogéniques ou iatrogènes dans la catégorie des hypertensions secondaires. Le niveau de granularité de la SNOMED-CT est parfois insuffisant. Il n'y a par exemple pas de distinction entre les douleurs de repos des membres inférieurs de nature ischémique et les autres, alors qu'elle conditionnent des attitudes diagnostiques et thérapeutiques différentes.

Enfin, la structure de la SNOMED-CT est inadéquate pour notre application, parce qu'elle ne reflète pas la façon dont les cliniciens pensent la prise en charge de l'hypertension. Par exemple, la SNOMED-CT ne comporte pas le concept structurant d'atteinte des organes cibles. Certains concepts sont même mal classés sur un plan strictement taxinomique. Le concept de nécrose ischémique devrait être dans la catégorie des signes cliniques (indiquant une maladie ischémique critique). Pourtant, le terme « gangrène ischémique » [402861007] n'est placé dans la SNOMED-CT que dans la catégorie des maladies et il n'y a pas de terme pour la donnée d'examen correspondante. En fait, la hiérarchie de la SNOMED-CT n'est même pas strictement taxinomique. Le terme « gangrène ischémique » est positionné comme un type de maladie artérielle, alors qu'il

ne s'agit pas d'une maladie de l'artère mais de la conséquence de l'insuffisance fonctionnelle artérielle.

L'inadéquation prévue des ontologies générales pour notre application et l'absence d'ontologie de domaine dédiée à la prise en charge de l'hypertension en milieu spécialisé nous ont conduit à entamer la construction d'une ontologie spécifique.

4.2 Nécessité d'une démarche de modélisation « sur mesure »

Notre démarche de modélisation ontologique se rattache à des réalisations antérieures qui s'appuient également sur l'analyse de gros corpus textuels produits en pratique clinique (Charlet *et al.*, 2006; Baneyx *et al.*, 2006). Ce cadre de modélisation assure que l'ontologie résultante contient les concepts qui sont effectivement utilisés par les cliniciens durant leur activité. Toutefois, nous avons adapté cette démarche à nos besoins : (i) sur le plan du matériel, en choisissant un corpus original, et (ii) sur le plan de la méthode, en exploitant également des conceptualisations expertes préexistantes du domaine.

Les corpus utilisés dans les réalisations antérieures étaient pour l'essentiel des collections de comptes rendus d'hospitalisation, alors que nous avons travaillé avec les observations cliniques. Dans la mesure où ces observations sont semi-structurées, nous avons pu directement en extraire un certain nombre de concepts importants. Savoir qu'un concept provient de la partie structurée de l'observation est un argument fort pour affirmer qu'il participe à une prise en charge réglée de l'hypertension artérielle, selon des règles explicites ou implicites en vigueur dans l'unité spécialisée. Nous avons complété cet ensemble de concepts relevant de la prise en charge réglée grâce à l'analyse des guides de bonne pratique. Nous avons également pris en compte les concepts d'occurrence fréquente dans les commentaires en texte libre, qui sont pour la plupart liés à des aspects routiniers de la prise en charge de l'hypertension. Ces trois strates réunissent les concepts fondamentaux utilisés pour la prise en charge réglée des problèmes de routine en hypertension artérielle. Il est probable qu'une première organisation des concepts sera plus simple à réaliser sur ce noyau de taille réduite et relativement autonome sur le plan ontologique, en ce qu'il représente complètement un des aspects de la prise en charge des patients.

La progression séquentielle permet donc d'identifier en premier lieu les concepts liés à la prise en charge réglée de l'hypertension. Par différence, nous devrions pouvoir facilement discerner les concepts liés à la pris en charge individualisée parmi ceux d'occurrence moins fréquente dans les commentaires en texte libre. Travailler avec les comptes rendus d'hospitalisation n'aurait pas donné les mêmes indications. Les guides de bonne pratique auraient été insuffisants pour indiquer le mode d'intervention des concepts dans la prise en charge de l'hypertension, car ils concernent essentiellement la prise en charge de l'hypertension en médecine générale et ne couvrent pas les aspects standardisés de la pratique en milieu spécialisé.

L'auteur principal disposait de connaissances cliniques suffisantes pour n'avoir recours aux experts du domaine que dans des cas problématiques. Lorsque les compétences sont séparées, une collaboration inter-disciplinaire plus étroite est nécessaire avec toutes les difficulté d'organisation et de communication qu'elle comporte.

4.3 Perspectives

Modélisation ontologique. Nous organisons actuellement le noyau ontologique en une monohiérarchie strictement taxinomique, dans le respect des principes de la sémantique différentielle. Nous commençons par structurer les concepts provenant des items de l'observation et des guides de bonne pratique, en nous inspirant des indices obtenus lors de l'analyse des ressources terminologiques (concepts structurants) et des relations taxinomiques trouvées dans la SNOMED-CT. Nous ajouterons ensuite les concepts récurrents dans les commentaires en texte libre, en réorganisant la structure si nécessaire. Une fois ce noyau ontologique consolidé, nous poursuivrons l'analyse terminologique du corpus de commentaires en texte libre, de manière à intégrer dans l'ontologie les concepts moins fréquents.

Analyse de l'individualisation des pratiques. Une fois l'ontologie construite, nous l'utiliserons pour représenter les cas cliniques pris en charge dans l'unité d'hypertension. Ll'association à chaque cas des concepts qui lui correspondent n'est complètement automatisable que pour la partie structurée de l'observation. En effet, les items d'observation sont liés de façon univoque avec les concepts qu'ils ont révélés. Par opposition, le contenu des commentaires en texte libre ne peut pas être automatiquement représenté avec les concepts qui en sont issus.

C'est donc d'abord sur la représentation automatique mais partielle des cas qu'une mesure de similitude sémantique opérera pour regrouper les cas similaires. L'analyse qualitative des cas similaires conduira à identifier des caractéristiques singularisantes importantes pour la décision médicale et permettra de compléter manuellement la représentation ontologique de chaque cas. À l'issue de cette première boucle, un nouveau calcul de similitude sémantique entre les représentations enrichies pourra définir des groupes de cas plus proches afin de raffiner l'analyse qualitative. Le processus pourra être répété autant de fois que nécessaire et possible. La mise en œuvre applicative de l'ontologie tiendra lieu de banc d'essai et de validation ultime de sa pertinence, en terme de pouvoir de représentation (étendue, granularité) et de calcul (structure).

5 Conclusion

L'analyse de l'individualisation des pratiques lors de la prise en charge de patients hypertendus dans une unité spécialisée nécessite une représentation détaillée du contenu sémantique des cas cliniques. Nous construisons une ontologie spécifique pour coder les cas médicaux de manière à la fois formelle et détaillée, afin de pouvoir nous aider d'outils informatiques pour l'analyse de l'individualisation des pratiques. Une démarche de modélisation originale par étapes successives a été définie afin d'optimiser l'adéquation de l'ontologie à la tâche visée.

Les items du formulaire d'observation utilisé dans l'unité d'hypertension ont été pris comme point de départ car ils reflètent une conceptualisation éprouvée de la prise en charge standardisée des patients en milieu spécialisé. Pour compléter ces concepts et obtenir des indices sur l'organisation ontologique, deux corpus ont été analysés. Chacun reflète un aspect supplémentaire des décisions cliniques : les guides de bonne pratique

correspondent à des règles de prise en charge validées pour la médecine générale et les commentaires en texte libre reflètent l'individualisation des décisions médicales pour des patients adressés dans une unité spécialisée. Grâce à cette démarche de construction, l'ontologie doit contenir tous les concepts utiles à l'évaluation des décisions médicales dans une unité d'hypertension, à la fois dans leur aspects réglés et dans leur aspects personnalisés.

Remerciements. Didier Bourigault a mis à notre disposition les résultats d'analyse des corpus par SYNTEX et UPÉRY. Ce travail bénéficie d'une subvention de la Société Française d'Hypertension.

Références

BANEYX A., CHARLET J. & JAULENT M.-C. (2006). Building an ontology of pulmonary diseases with natural language processing tools using textual corpora. *Int J Med Inform*, **76**, 208–215.

BOURIGAULT D. (2002). Upéry : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *9e conférence annuelle sur le traitement automatique des langues*, p. 75–84, Nancy.

CHARLET J., BACHIMONT B. & JAULENT M.-C. (2006). Building medical ontologies by terminology extraction from texts : an experiment for the intensive care units. *Comput Biol Med*, **36**(7-8), 857–870.

COIERA E. (1995). Medical informatics. *BMJ*, **310**(6991), 1381–1387.

DEGOULET P., CHATELLIER G., DEVRIÈS C., LAVRIL M. & MÉNARD J. (1990). Computer-assisted techniques for evaluation and treatment of hypertensive patients. *Am J Hypertens*, **3**(2), 156–163.

FRIEDMAN C. & WYATT J. (2006). *Subjectivist approaches to evaluation*, In W. J. FRIEDMAN, C.P., Ed., *Evaluation methods in biomedical informatics*, chapter 9, p. 248–266. Health informatics series. Springer : New-York, 2nd edition.

GREEN J. & BRITTEN N. (1998). Qualitative research and evidence based medicine. *BMJ*, **316**(7139), 1230–1232.

HAYNES R., DEVEREAUX P. & GUYATT G. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *Evid Based Med*, **7**(2), 36–38.

STEICHEN O., DANIEL-LE BOZEC C., THIEU M., ZAPLETAL E. & JAULENT M.-C. (2006). Computation of semantic similarity within an ontology of breast pathology to assist inter-observer consensus. *Comput Biol Med*, **36**(7-8), 768–788.

TRONCY R. Differential ontology editor. `http://homepages.cwi.nl/~troncy/DOE/`.

VIRGINIA-MARYLAND REGIONAL COLLEGE OF VETERINARY MEDICINE. SNOMED-CT browser. `http://snomed.vetmed.vt.edu/sct/menu.cfm`.

WANG A., SABLE J. & SPACKMAN K. (2002). The SNOMED clinical terms development process : refinement and analysis of content. *Proc AMIA Symp*, p. 845–849.